1-1-2012

Conceptual Problems Of Visual Literacy

Paul Messaris
University of Pennsylvania, pmessaris@asc.upenn.edu

Conceptual Problems Of Visual Literacy

2

CONCEPTUAL PROBLEMS OF VISUAL LITERACY

Paul Messaris

In one of the foundational texts of visual scholarship, *Studies in Iconology*, distinguished art historian Erwin Panofsky (1939) provides what is probably the most complete account to date of how many different layers of meaning can be present in a single picture. In the process of parsing the historical roots and cultural resonances of such traditional images as father time or the winged cupid, Panofsky gives a stunning demonstration of the amount of erudition that is entailed in an informed reading of pictures that may, on the surface, appear quite simple. Anyone who has read Panofsky's book will have come away with a renewed appreciation for the complexity of this form of erudition, which fairly may be labeled *visual literacy*. Yet Panofsky himself raises the issue of another kind of visual interpretation whose dependence on prior literacy is by no means as clear. As Panofsky indicates, we can make a distinction between two different aspects of pictorial interpretation: on the one hand, understanding an image's cultural significance; on the other hand, perceiving the image's physical contents—the people, places, and events depicted in it. It is self-evidently true that the former aspect of interpretation requires visual literacy. But what about the latter? As a prelude to answering this question, let us examine an act of spectatorship that is far removed from the urban European context in which Panofsky was writing.

In a magazine article with major implications not only for visual communication but also for verbal language, reporter John Colapinto describes a visit to the rain forest of northwestern Brazil, where he spent some time

with members of an Amazonian tribe called the Piraha, consisting of approximately 350 people. Colapinto was accompanying a U.S. linguistics professor, John Everett, who has been studying the Piraha for some 30 years and is one of very few outsiders with any substantial knowledge of their language and culture. The language of the Piraha is a subject of major controversy in the linguistic community because its structure is so different from that of any other known language as to call into question certain fundamental assumptions about human linguistic universals. In particular, Everett observes that Piraha grammar confines its speakers to the discussion of nonabstract entities within their immediate experience, and therefore lacks such things as numbers (or even the concept of counting), color terms, embedded clauses, and relative tenses. Moreover, Piraha culture is notable for "the absence of creation myths and fiction, the absence of any individual or collective memory of more than two generations past, the absence of drawing or other art" (Everett, 2005, p. 621). A couple of days into his visit, Colapinto's impression of the Piraha was one of unbridgeable cultural remoteness.

Then one evening, Colapinto attended a film screening that Everett had arranged for an audience of about 30 Piraha with the aid of a generator-powered DVD player. The movie was the 2005 Peter Jackson version of *King Kong*, and the viewers' reactions were enthusiastic. "The Piraha shouted with delight, fear, laughter, and surprise—and when Kong himself arrived, smashing through the palm trees, pandemonium ensued" (Colapinto, 2007, p. 136). Moreover, the Piraha's comments also made it clear that they fully understood the implications of the less action-oriented scenes in which King Kong interacts with the character played by actress Naomi Watts. Despite the fact that these scenes involve radical disruptions of the continuity of space and perspective as the camera cuts back and forth between huge close-ups of the two protagonists, the Piraha appeared to have no trouble interpreting the scenes' implied sexual undercurrents. The ease with which the Piraha viewers interpreted this particular movie becomes even more remarkable when one takes into account the fact that their prior exposure to cinema had been confined entirely to Everett's personal collection. Their environment contains no movie theaters, no television, no Web access, and, indeed, no regular supply of electricity. As Colapinto (2007) concludes, in contrast to research findings on Piraha language and cognition, "Jackson's movie left no question about the universality of Hollywood film grammar" (pp. 136-137). Indeed, Everett (2005) pointedly observed that, despite more than 200 years of regular contact with neighboring Brazilian peoples, the Piraha do not speak any verbal language other than their own.

This story is a vivid illustration of just how different pictures are from verbal language. People who study visual communication customarily talk

about the need for visual literacy. As a convenient label for pictorial competence (ability to create images, interpret them adequately, discern their consequences, resist their effects), the term is certainly useful, and, in any case, there is no readily available substitute. However, if the term carries with it the implication that pictorial fluency works in the same way as verbal fluency does, its use is substantially misleading. It leads to the assumption that the interpretation of pictures depends on prior familiarity with a set of formal conventions, whereas one of the essential characteristics of pictorial communication is precisely the fact that it can take place even when those conventions are unfamiliar or are encountered for the first time, as in the case of the Piraha. A popular textbook on visual communication makes the following claim: "There are certain codes or conventions of [movie] editing that we all learn from experience, from watching numerous motion pictures. ... All of these devices are analogous to punctuation in writing. ... Someone unfamiliar with these editing conventions, is likely to misinterpret the film" (Berger, 2008, p. 151). It is because of this kind of overstated analogy between pictures and language that the term *visual literacy* needs to be treated with caution, if not discarded altogether. To put this differently: Meyrowitz (1998) argues that literacy in any medium entails three types of knowledge, having to do with the medium's contents, its fixed characteristics, and the production variables that are employed by its users. Meyrowitz calls this third type of knowledge "media grammar literacy." Using his terminology, we can say that a distinguishing feature of visual media is that they are the one form of communication whose competent interpretation does not depend on media grammar literacy.

If the role of literacy (i.e., prior familiarity with conventions) is not as pivotal in pictorial interpretation as it is in linguistic comprehension, then it must be the case that, for some visual conventions at least, certain general cognitive processes (i.e., mental operations whose application extends beyond the range of pictures and that a viewer could be expected to develop even before his/her first encounter with pictorial media) provide an adequate basis for interpretation. My goal in this chapter is to spell out what these cognitive processes might be. Specifically, I discuss the following four types, which I consider to be central components of a competent viewer's repertory: (a) the brain's processing of visual information in a modular fashion; (b) analogical thinking—the ability to perceive a formal analogy between a visual device and some aspect of everyday experience; (c) spatial intelligence—the derivation of a coherent sense of a three-dimensional scene out of a limited number of partial views of that scene; and (d) the use of contextual information in the interpretation of nonverbal behavior. In discussing these cognitive processes, I develop a theoretical account of visual interpretation that differs from the more tra-

ditional, visual literacy approach. This is not to say that the two approaches are totally incompatible. In fact, I describe ways in which they might complement each other. However, to the extent that the approach I suggest offers a valid interpretation of the viewing process, one of the main implications of the visual literacy approach—the idea that pictorial comprehension requires prior exposure to images—needs to be rethought.

MODULAR PROCESSING OF VISUAL INFORMATION

A major impetus behind the early development of the visual literacy idea was the work of art historian E.H. Gombrich (2004)—in particular, his book on *Art and Illusion*, whose first edition was published in 1960. Arguing against the "commonsensical" view that pictorial comprehension is based on simple similarity between pictures and the things they represent, Gombrich pointed out that there are numerous ways in which even the most "realistic" picture differs from the appearance of the real world. For example, in the real world, objects have three dimensions, whereas, on a picture surface, the world is flat. Moreover, many kinds of pictures that we may accept unthinkingly as adequate representations of reality are actually quite radical departures from it. A black-and-white movie exhibits many of the attributes of "photographic realism"—but it lacks color. A technical blueprint is a mathematically accurate representation of spatial relationships—but it lacks shading as well as color. A cartoon creature may be beguilingly real to audiences of all ages—but it lacks accurate proportions. Gombrich's writings contain many such examples of sharp discrepancies between the world of visual images and the world of three-dimensional, naturally proportioned, physically illuminated reality. In the face of such discrepancies, the concept of visual literacy may seem to offer a self-evidently true account of how people come to comprehend images: They have to learn the conventions, because so many of the conventions are "unrealistic."

Gombrich's influence on the thinking of subsequent generations of visual scholars has been long lasting, and, as we have seen, the visual literacy idea is still present in textbooks on visual communication. However, much of the empirical evidence on picture perception is at odds with this idea. Several cross-cultural studies (see Messaris, 1994, for a review), as well as some research with animals (Fagot, 2001), and even a study of blind people's interpretations of raised outline drawings (Kennedy, 1993) have found that a variety of pictures that may intuitively appear "unrealistic" are actually understood without much difficulty by untutored viewers. Perhaps the most noteworthy single study in this area

is a classic, one-of-a-kind experiment by Hochberg and Brooks (1962), husband-and-wife scholars who raised their own child without any exposure to pictures and then tested his picture-perception abilities when he was old enough to be able to talk (see also Peterson, Gilliam, & Hedgwick, 2006). The experiment contained not only black-and-white photographs but also unshaded outline drawings (i.e., two styles of pictures that both contain significant deviations from the appearance of unmediated reality). And yet, the child in this experiment was able to give correct identifications of the contents of almost all of the pictures he was shown.

What cognitive process(es) could account for such findings? One possible way to explain them comes from the theory of modularity of visual perception (see Calabretta & Parisi, 2005; Viviani & Aymoz, 2001; Zecki, 2005). According to this theory, our perceptual system encompasses a variety of distinct processes or "modules" that interpret the various components of a visual scene—shapes, movements, colors, and so on—separately from one another. In other words, the cognitive process that determines what shapes we are looking at functions separately from the cognitive process that detects and interprets the nature of movement, and so forth. The crucial point here is that, in theory, the operations of certain modules are independent of the operations of other modules. In principle, we might be able to tell what types of movements we were witnessing even if for some reason we weren't able to say what types of objects were performing those movements. Paradoxical although this supposition may seem, it has actually received considerable substantiation from studies of patients with pinpoint brain damage that affects one or more of the perceptual "modules" while leaving the others intact. One of the most dramatic examples of this kind of thing was described by neurologist Oliver Sacks (1998) in his widely publicized study, "The Man Who Mistook His Wife for a Hat." The patient in this case had a brain malfunction that robbed him of the ability to identify objects, but he was still perfectly able to describe their shape, size, movement, and so on.

How does the theory of modularity of vision apply to picture perception? Perhaps the simplest example of its implications is the problem of pictorial depth perception. Except for so-called "3D" images that are viewed with special glasses, all pictures (including movies) lack a crucial item of information about distance or "depth"—namely, binocular disparity, the difference between what the left eye sees and what the right eye sees. This difference is an extremely powerful depth cue, especially for objects that are close to us. If vision functioned "wholistically" and everything had to be in place simultaneously for us to be able to make sense of our visual impressions, the fact that pictures lack binocular disparity would be an insurmountable obstacle for first-time viewers of pictures, who would need to learn how to interpret pictures in its absence (i.e.,

they would have to acquire visual literacy). However, according to the assumptions of the modular theory, lack of binocular disparity should not prevent inexperienced viewers from interpreting pictures if other types of visual information are present in the picture and are capable of sustaining an adequate interpretation. In the specific case of depth perception, there is at least one other very powerful depth cue (viz. motion parallax) that is always present in motion pictures, and that should provide a sufficient sense of depth in and of itself. When King Kong comes crashing through the jungle, it is mainly motion parallax that tells us he is getting closer, and even inexperienced viewers should be able to make that inference—as indeed happened with the Piraha audience described by Colapinto.

Modularity of vision also is almost certainly a major factor in our ability to make sense of black-and-white photographs, as well as other types of images that lack naturalistic color. There is considerable evidence that color perception is largely localized in certain specific areas of the brain, and there is also evidence that when these areas malfunction and the color "module" cease to operate normally, the other aspects of vision are able to proceed without significant impediment. A real-world example of this kind of situation has been described by Sacks in another of his well-known case studies, "The Painter Who Became Color Blind" (Sacks & Wasserman, 1987), about an artist who lost his ability to see color as a result of an accident but was still able to make sense of all other features of the visual field (depth, motion, object identity, etc.). By extension, it should not surprise us that the young child studied by Hochberg and Brooks was able to make sense of black-and-white photographs the very first time he encountered them. The child's ease of interpretation of outline drawings is also partly explainable on the basis of the modularity of color vision, but a full explanation of that aspect of Hochberg's and Brook's findings (as well as similar findings in studies of animals) is somewhat more complicated and would take us beyond the scope of this chapter. Suffice it to say that there is good reason to believe that our brain's processing of real-world visual information entails the extraction of "outlines" of the objects in our visual field, and that this process most likely provides part of the "module" that is responsible for our ability to make sense of sketches, stick figures, and other kinds of incomplete pictorial representations (Marr, 1983). The more fundamental point in all of this is that, if the modular theory is even partly correct, it furnishes an explanation of pictorial perception that does not depend on prior familiarity with pictures. In other words, according to this theory, we are able to make sense of pictures because they elicit cognitive mechanisms that we already use in everyday visual perception—even if we have never seen a picture before and are therefore totally "illiterate" in that respect.

ANALOGICAL THINKING

One of the clearest examples of a formal convention that is encountered very widely both in still and in moving pictures is the use of a low or a high camera angle as a means of making someone look powerful or powerless. I would assume that even someone with no formal background in visual scholarship would readily recognize this convention because it is, as indicated, in very wide use. However, for our purposes the important question is this: How does the viewer come to understand this convention when he or she sees it in a print ad or film or TV program? What previous knowledge must the viewer have in order to be able to respond to this use of camera angle in the appropriate manner (as called for by the convention)? The visual literacy response to this question would be that the viewer would have to have had a number of previous encounters with this use of camera angle, in the course of which she or he would gradually have acquired a sense of this device's meaning. For example, we may imagine a child gradually coming to associate low angles with shots of villains in threatening postures and thereafter responding to the angle itself in the appropriate way. This certainly seems like a plausible possibility, and, more than likely, it does indeed happen to a certain extent.

At the same time, however, it seems to me that there is an alternative route to the interpretation of this kind of use of camera angle—and, if I am correct, this alternative route would not require any previous exposure to this specific device. This alternative route is based on the fact that the camera-angle convention is not an arbitrary convention (in other words, it's not like the word "powerful," whose form is unrelated to the concept it denotes); instead, I would argue that the particular use of camera angle which we've been examining derives its meaning by analogy with real-life situations of looking up at powerful people or looking down at weak people—a realm of experience that is likely to be particularly relevant to the formative years of childhood. If this assumption is correct, its implication is that a viewer should be able to respond appropriately to camera angle on the basis of the analogy with real-life experience, without any necessary previous exposure to the use of camera angle in pictorial media. In other words, here we have another example of a pictorial element that may be interpretable on the basis of a general cognitive process, namely, sensitivity to visual analogy—or, perhaps, even more generally, analogical thinking. This alternative possibility does not preclude the visual literacy approach, and, in fact, it seems quite possible that general sensitivity to visual analogy may develop from the specific experience of camera angle and other similar pictorial conventions.

Nonetheless, as interpretational mechanisms, these two alternatives are certainly distinct.

Although camera angle may be one of the clearest examples of a device that draws on analogy for its meaning, in my view it is certainly not the only one. Indeed, I argue that the use of analogical constructions is one of the distinctive features of still and moving images as modes of communication. Let me list, very briefly, some other formal conventions whose meaning appears to derive from analogy with some aspect of real-life experience. At a bare minimum, such a list would include the following:

1. Camera-to-subject distance (i.e., the use of close-up vs. medium shot, etc.) as a means of emphasis or as a means of generating intimacy and identification with a character on the screen (as Reeves & Nass, 2003, have argued in an extended analysis of this variable, it appears to derive its meaning and effectiveness from an analogy to the real-life area of proxemic behavior).
2. The use of camera movement to simulate a character's subjective visual experience in point-of-view shots (Farrar, Krcmar, & Nowak, 2006)—for example, the "stealthy" camera movement that tells us we are looking through the eyes of the serial killer in a slasher movie; and
3. Speed of editing as a means of modulating the emotional tone of a scene—for example, the use of rapid cutting to make action scenes more exciting, or the use of slow-paced editing to made romantic scenes more languorous.

SPATIAL INTELLIGENCE

As the list just presented indicates, the scope for analogical thinking in the interpretation of visual media appears to be quite extensive. However, in my view, the cognitive process which is of greatest importance to film or TV editing, in particular, is not analogical thinking but, rather, spatial intelligence. As conceived of by cognitive psychologists (e.g., see Gardner, 1999), spatial intelligence comprises a cluster of related cognitive abilities, of which the most crucial for our purposes is the ability to derive a coherent sense of a three-dimensional scene out of a limited number of partial views of that scene. Anyone familiar with cognitive psychology will recognize here an area of intelligence that is typically tapped through such measures as Piaget's three-mountain task: A child is shown a certain view of a mountainous landscape and asked to indicate how the mountains

would appear from a different viewing position. Although I do not think that this specific situation has an exact parallel in film or TV interpretation, the general mental process of spatial integration on the basis of partial views is brought into play every time the action in a scene is "interrupted" by a cut from one point of view to another. Of course, such transitions need not be extreme. Often, all that is involved is a small reorientation of the camera back and forth between two people having a conversation. On the other hand, when it comes to action sequences, or such things as a switch from an "objective" view to a "subjective" shot (i.e., the point of view of a character in the scene itself) the change in point of view can be quite radical—and, presumably, quite demanding with regard to the viewer's spatial intelligence.

As I have indicated, I think that spatial intelligence may be the most important component of a competent viewer's repertory of cognitive processes for interpreting editing. This judgment is based on an analysis of the kinds of editing that a viewer is likely to encounter in typical fictional TV programs. In an unpublished study, my students and I examined a convenience sample of nine U.S. TV programs: three daytime soap operas, three sitcoms, and three crime dramas. Our analysis was concerned with the editing. We looked at each shot transition (cut, fade-out/fade-in, dissolve, etc.) and classified them into five overall categories, of which the only one that is relevant for our purposes was the first: a transition within a single location, from one point of view to another. Overall, an average of 95% of the transitions fell into this category. (Average *N* for total transitions, 559 per soap opera; 250 per sitcom; 398 per crime drama.) In other words, by an overwhelming majority, the kind of editing transition that a viewer is likely to be confronted with in a typical fictional TV program (at least in the United States) is precisely the kind of transition for which spatial intelligence is the relevant interpretational process. All the other editing devices—time-space changes, flashbacks, and so on, which sometimes seem to get the lion's share of attention from scholars—are in fact a tiny minority of the whole.

The relevance of spatial intelligence to film or TV viewing has received considerable attention from cognitive psychologists, and there are several studies suggesting or demonstrating a link between TV experience and performance on Piagetian or other tests of this cognitive process (most importantly, Tidhar, 1984). A review of this research is beyond the scope of this chapter, but the general finding—namely, that TV viewing can influence spatial intelligence—suggests a two-sided conclusion to what has been said so far: On the one hand, the major thrust of this presentation has been to argue that, when certain cognitive abilities precede visual-media experience, they can provide an avenue to interpretation in the absence of specific familiarity with the formal conventions of these

media. On the other hand, to the extent that spatial intelligence—and perhaps other cognitive processes—are developed further through the viewing experience itself, we could say that competence in film or TV interpretation is actually a form of more general intelligence.

USE OF CONTEXTUAL CUES IN NONVERBAL COMMUNICATION

The creation of a coherent space–time continuum out of the fragments presented in a film or TV program is one of the central intellectual tasks that visual media demand of their viewers. However, the point of editing is not always that of linking time frames and points of view. A second major purpose—especially in dramatic contexts—is that of revealing characters' thoughts, intentions, and personalities. In his description of the Piraha audience's reactions to *King Kong*, Colapinto (2007) makes a point of the fact that their verbal commentary contained explications of "what was being communicated in the long, lingering looks that passes between gorilla and girl" (p. 137). The use of editing to imply thoughts and feelings was one of the earliest discoveries in the history of explicit theorizing about the movies. Its formulation is usually associated with Lev Kuleshov and other filmmakers working during the early years of Soviet cinema. In its best-known incarnation, the so-called "Kuleshov effect" is illustrated in Kuleshov's experiment involving an "expressionless" close-up of the Russian actor Mozhukhin juxtaposed with a variety of other scenes, including a plate of soup on a table, a corpse in a coffin, and a little girl playing with a toy bear. According to Kuleshov's colleague V.I. Pudovkin (1976), to whom we owe the best-known description of this experiment, viewers who saw these sequences without having been told about the editing responded with enthusiastic praise for Mozhukhin's acting. In other words, the editing led these viewers to see subtle changes of expression—from thoughts of food to deep sorrow to a "light, happy smile"—where in fact there were none (p. 168).

The general category of juxtaposition explored in this experiment (and others that followed it) is a firmly established feature of film and TV editing, occurring most commonly, perhaps, in the conventional "reaction-shot" sequence, in which shots of a speaker or other objects of interest are intercut with shots of a listener or observer. Such sequences are a typical ingredient of dialogue scenes in fiction films, as well as of "nonfictional" dialogues in talk shows or other TV programs, but the potential role of image juxtaposition as an indicator of characters' thoughts or reactions is probably most evident in the absence of dialogue and in those "nonfiction-

al" cases in which a certain sequence of events is rearranged through editing (as in the many instances in which an interviewer's "reactions" are inserted into a TV interview after the fact). Assuming, as the evidence suggests, that viewers typically do use the juxtaposition of images—rather than just the facial expressions in them—as clues to what lies "beneath the surface" of characters' faces, we are confronted with another cluster of visual conventions based on a single general principle. What might account for viewers' ability to make sense of these conventions?

One possibility, as always, is that entailed in the notion of visual literacy, namely, previous experience with the conventions in question. On the other hand, this is an area in which a ready parallel with a set of real-life cognitive processes suggests itself. Although the precise visual sequences that the viewer is confronted with on the screen—a view of a character juxtaposed with a view of some object or situation of interest to that character—may not have an exact parallel in reality, the basic inferential task that the viewer has to perform in the case of the film or TV sequence is similar to an extremely common real-life task, namely, that of judging other people's intentions from the context of their behavior. The degree to which this process is central to interpersonal communication bears some emphasis. As researchers in the area of nonverbal communication and of person perception have noted, people's appearance, expressions, and actions are frequently ambiguous, or even completely opaque, in the absence of information about the objects or situations to which they are addressed. Indeed, a classic theoretical treatise by Birdwhistell (1970) has advanced the argument that no facial expression or gesture has a determinate meaning out of context. The ability to take context into account in inferring thoughts and assessing intentions is consequently a vital component of any mature person's social skills. It is conceivable that this ability—rather than any direct experience with editing conventions—may serve as the basis of the interpretational competence called for by the kinds of editing we are concerned with here.

This notion—that the "Kuleshov effect" and related cinematic devices are derivatives of the real-life dependence of meaning on context—is consistent with the implications of another, less widely known, experiment by the Russian filmmaker. In this experiment, Kuleshov filmed an actor in two roles: first, in a jail cell, as a famished prisoner being offered a bowl of soup; second, as a prisoner released from jail and taken out into the open air. The actor was invited to use every means at his disposal to express the sentiments appropriate to these two situations: on the one hand, craving for the soup; on the other hand, delight at the sight of birds, clouds, the sun. Then Kuleshov (1974) produced various versions of the two scenes, in some of which the shots of the actor were transposed from one scene to the other. By his own account, regardless of how the scenes

were scrambled, viewers were unable to detect any discrepancy in the actor's performance.

In other words, despite the fact that the actor had a clear and distinct sentiment in mind in each case, his facial expressions in themselves were apparently incapable of conveying a specific enough sense of his thoughts, and the viewers' ultimate interpretations were evidently fixed by the overall context. This is essentially the point that investigators of real-life social perception have made about the information available in facial expressions and other overt indicators of thought and intention. Unlike the more famous Kuleshov experiment mentioned earlier, whose use of an unvarying, neutral expression might be seen as somewhat artificial, this one is based on a closer approximation of real-life conditions, in the sense that the actor's performance was allowed to vary with the situation, and it therefore makes the potential relationship between real life and this aspect of movie viewing clearer.

CONCLUSION

As this discussion suggests, then, both real-life experience and exposure to visual media are potential avenues to the interpretation of the kinds of visual devices we have been considering. One possibility does not necessarily exclude the other, of course. It is conceivable that these two sources of interpretational competence might work together, either by reinforcing each other or by interacting in a more complex fashion. For example, previous exposure to editing might teach a viewer which juxtapositions of images to look at for psychological implications, while real-life social experiences might guide the actual inferences drawn from those juxtapositions. In the absence of research aimed explicitly at disentangling these possibilities—for example, a study of "naïve" adult viewers' susceptibility to the Kuleshov effect—it is unclear that one can be more specific about either the necessary preconditions of these aspects of visual interpretation or the typical mix of experiences leading to them. However, what we can say is that, to the extent that the connections that have been drawn here between general (real-life) cognitive processes and visual conventions are valid, the interpretation of these conventions should be accessible even to an inexperienced, first-time viewer. In short, the assumption that interpretation requires visual literacy may be wrong.

If these beliefs about visual literacy are indeed mistaken, they can be said to have done a considerable disservice to visual scholarship. They have obscured one of the most distinctive and consequential attributes of images as a mode of communication—namely, the fact that images,

unlike words, engage our eyes and brains in the same way that the real word engages them. As we have seen, even when images are superficially "unrealistic" (black-and-white photographs, sketches, etc.), they are still capable of evoking the mental processes that we use in our transactions with raw, unmediated reality. This feature of images gives them a kind of power that words can never have. Understanding that power in all its ramifications should be a central goal of visual scholarship. Our pursuit of that goal is impeded by faulty analogies between images and words, and by inappropriate assumptions regarding visual literacy.

What are the educational implications of this conclusion? Does all this mean that educators can safely ignore this aspect of visual literacy—as opposed to the kind of cultural education about images that Erwin Panofsky was concerned with? Quite the contrary. The conclusion that we have arrived at contains a paradox: Precisely because the basic language of visual communication can be so transparent, learning to deal critically with that language may actually be more of a challenge than the critical parsing of verbal texts. For example, a low camera angle that tries to make a politician appear powerful may "fly under the radar" of citizens' perceptions more effectively than the verbal proclamation, "Candidate X is a strong leader." As an attempt to discourage the uncritical acceptance of such visual devices, education in visual literacy may be very valuable indeed.

In other words, if a major goal of education is to encourage learners to think for themselves, the main focus of visual education should not be on acquiring the grammar of images as such. Young people already have an intuitive understanding of that grammar. Instead, the central task should be to foster a conscious and critical *awareness* of the grammar, even when it may appear so simple as to be self-evident. This point deserves some emphasis. Anyone who has taught visual communication is familiar with the fact that the most powerful means of manipulating viewers are also, very often, the simplest—and that, in consequence, students often have very little appreciation of that power unless it is deliberately brought to their attention. Consider the visual principle of "false continuity," which is one of the most fundamental premises of the illusionistic power of film and television. Essentially, it is the basic principle behind most narrative editing: two shots joined together in the context of a broader narrative are "read" by the viewer as being part of a coherent stream of space, time, and action—even if the shots were in fact taken at widely separate times and places or if the actions within them were completely unrelated in real life (Deren, 2005). This principle is equally a part of fictional and nonfictional narratives, but it is the nonfictional case in particular that raises troublesome questions of visual manipulation and the need for critical viewing.

In its most extreme form, this principle may be observed in operation in those situations in which shots of nonfictional events are assembled after the fact for inclusion in a documentary, newscast, interview program, and so on. A notable example of the kind of misrepresentation that such a situation can lead to occurred in a July 2007 BBC report of an encounter between England's Queen Elizabeth and Annie Leibovitz, a U.S. portrait photographer. In preparation for a photographic session, Leibovitz reportedly asked the queen to remove her tiara, a suggestion that was apparently received with some initial asperity but that did not cause any major problems. However, in an edited tape of the event, the interaction with Leibovitz was followed by a shot of the queen expressing considerable displeasure—an incident that had actually occurred earlier, in reaction not to Leibovitz but to the heavy cape that the queen was wearing before the shoot. The BBC itself caught this error, and openly acknowledged its significance as a demonstration of the potentially misleading uses of editing.

Such scruples are less evident in other realms of image making, such as political advertising. It is a common procedure in political ads to portray the candidate addressing an audience of admiring representatives of the public. The intention, of course, is to convey an image of spontaneous approbation of the candidate's remarks. Often, however, the conjunction of candidate and responding public may be more of a product of editing than of the real-life encounter between the two. For example, in more than one ad used during the 2007-2008 election cycle in the United States, a rousing speech by a candidate is intercut with shots of a wildly enthusiastic audience—but, when one examines these images carefully, it becomes clear that the candidate is actually delivering his or her speech to a much smaller group of people than the large, cheering crowd that appears in the reaction shots.

Judging from own experience with students, misleading uses of such editing juxtapositions are very hard to detect on first viewing. The tendency to succumb to the illusion of false continuity is evidently very strong. Unfortunately, there is no systematic research that I am aware of regarding viewers' responses to this type of editing device, although the informal experiments described by Lev Kuleshov (1974) and other early Soviet filmmakers are certainly very instructive. Nevertheless, I strongly suspect that, as far as critical viewing is concerned, this aspect of visual manipulation may pose formidable challenges even to the literate viewer. And yet, the editing device itself is exceedingly simple: two shots joined together in the absence of an actual spatial or temporal connection between their contents. If our goal was to teach students to understand the intended meaning of the juxtaposition, or even to create such a sequence for themselves, the lesson would be over very soon. But developing the habits of

mind that enable people to spot such misleading juxtapositions when they are used by others is a more gradual undertaking. The need for that kind of visual literacy is the ultimate conclusion of the arguments developed in this chapter.

REFERENCES

Berger, A. A. (2008). *Seeing is believing: An introduction to visual communication* (3rd ed.). New York: McGraw-Hill.

Birdwhistell, R. L. (1970). *Kinesics and context: Essays on body motion communication*. Philadelphia: University of Pennsylvania Press.

Calabretta, R., & Parisi, D. (2005). Evolutionary connectionism and mind/brain modularity. In W. Callebaut & D. Rasskin-Gutman (Eds.), *Modularity: Understanding the development and evolution of complex natural systems* (pp. 309-330). Cambridge, MA: MIT Press.

Colapinto, J. (2007, April 16). The interpreter: Has a remote Amazonian tribe upended our understanding of language? *The New Yorker*, pp. 118-137.

Deren, M. (2005). *Essential Deren: Collected writings on film* (B. R. McPherson, Ed.). New York: Documentext.

Everett, D. (2005). Cultural constraints on grammar and cognition in Piraha: Another look at the design features of human language. *Current Anthropology, 46*(4), 621-646.

Fagot, J. (Ed.). (2001). *Picture perception in animals*. London: Psychology Press.

Farrar, K. M., Krcmar, M., & Nowak, K. L. (2006). Contextual features of violent video games, mental models, and aggression. *Journal of Communication, 56*(2), 387-405.

Gardner, H. (1999). *Intelligence reframed: Multiple intelligences for the 21st century*. New York: Basic Books.

Gombrich, E. H. (2004). *Art and illusion: A study in the psychology of pictorial representation* (6th ed.). London: Phaidon Press.

Hochberg, J., & Brooks, V. (1962). Pictorial recognition as an unlearned ability: A study of one child's performance. *American Journal of Psychology, 75*, 624-628.

Kennedy, J. M. (1993). *Drawings and the blind: Pictures to touch*. New Haven, CT: Yale University Press.

Kuleshov, L. (1974). *Kuleshov on film* (R. Levaco, Trans. & Ed.). Berkeley: University of California Press.

Marr, D. (1983). *Vision: A computational investigation into the human representation and processing of visual information*. New York: W.H. Freeman.

Messaris, P. (1994). *Visual "literacy": Image, mind, and reality*. Boulder, CO: Westview Press.

Meyrowitz, J. (1998). Multiple media literacies. *Journal of Communication, 46*(1), 96-108.

Panofsky, E. (1939). *Studies in iconology: Humanistic themes in the art of the Renaissance*. Oxford, UK: Oxford University Press.

Peterson, M. A., Gilliam, B., & Hedgwick, H. A. (Eds.). (2006). *In the mind's eye: Julian Hochberg on the perception of pictures, films, and the world.* New York: Oxford University Press.

Pudovkin, V. I. (1976). *Film technique and film acting* (I. Montagu, Trans. & Ed.). New York: Grove Press.

Reeves, B., & Nass, C. (2003). *The media equation: How people treat computers, television, and new media like real people and places.* Chicago: University of Chicago Press.

Sacks, O. (1998). *The man who mistook his wife for a hat.* New York: Touchstone.

Sacks, O., & Wasserman, R. L. (1987). The painter who became color blind. *New York Review of Books, 34,* 25-33.

Tidhar, C. E. (1984). Children communicating in cinematic codes: Effects on cognitive skills. *Journal of Educational Psychology, 76*(5), 957-965.

Viviani, P., & Aymoz, C. (2001). Color, form and movement are not perceived simultaneously. *Vision Research, 41*(22), 2909–2918.

Zecki, S. (2005). Behind the seen: The functional specialization of the brain in space and time. *Philosophical Transactions of the Royal Society London B: Biological Sciences, 360,* 1145-1183.